Burgess, F.M. The validation study of the Threading Sleeves test.
1947

BOSTON UNIVERSITY

SCHOOL OF EDUCATION

Thesis

THE VALIDATION STUDY OF THE THREADING SLEEVES TEST

Submitted by

Faye M. Burgess

B. S. in West Texas State Teachers College, 1935

In partial fulfillment of requirements for the degree of Master of Education

1947

First Reader: William C. Kvaraceus, Assistant Professor of Education

Second Reader: John L. Rowe, Assistant Professor of Commercial Education

Third Reader: Dugall Arbuckle, Assistant Professor of Education

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION AND REVIEW OF PREVIOUS RESEARCH

INTRODUCTION OF THE PROBLEM

Statement of problem. The problem chosen for this study is to find out how accurately the Threading Sleeves Test can be used as an aid in selecting operators with the ability to become efficient mount operators within an electronic tube factory. The Threading Sleeves Test was one of a battery of tests that was made up for selecting operators with the ability to handle small parts with a steady hand and eye-hand coordination.

The development of the problem. The development of this test was the outgrowth of research and experiment within an electronic tube factory, Raytheon Manufacturing Company, Newton Massachusetts, during the war. It was at such a time that industry was expected to meet the demands of increased production. Its program was accelerated; there was a big job to do. New ideas and methods are more readily accepted than in ordinary times. The problem of selecting, placing, and transfering became a great concern of the company. Dr. Guy M. Wilson directed the development of the testing program and supervised the writer for two and one half years. It was at this time the Threading Sleeves Test was constructed to be used with a battery of tests chosen to help select operators for the Mount Department. The largest percent of all the employees hired were sent

to this department to be trained as mount operators. These operators worked together on teams. It was the usual procedure to place a fast worker on each team to set the pace for the group regardless of their ability for speed. Later research showed the value of selecting operators for a team with generally the same speed levels. That is, let the good operators work together on the same team, the average operators work together, etc. As this idea was gradually accepted, the need for tests results as a gauge became helpful. By careful guidance directed by Dr. Wilson, the supervisors were taught the wise use of these tests results. It was at this stage of instructing the supervisors that the problem of this paper became apparent. One supervisor mentioned the value of the Threading Sleeves Test in that she could use it solely to place operators on her teams. Even though the general results seemed favorable, no validation study had been made on this test.

Just how reliable is the Threading Sleeves Test? Could it be used as an aid in selecting mount operators? Before an attempt will be made to answer these questions, a general review of related studies will be investigated to see what has been reported on other dexterity tests.

<u>What are dexterity tests?</u> "Dexterity tests measures the ability to work rapidly and skillfully with the fingers, hands, and arms."[1] Certain jobs require rapid and skillful manipulation of materials. Attempts have been made to build instruments that will measure this

---

[1] Remmers, H. H. and Gage, N. L. <u>Educational Measurement and Evaluation</u>, Harpers Brothers, New York, 1943, p. 327.

skill on particular jobs. These tools are being examined more critically to a greater extent than heretofore. To follow this development of dexterity testing, it is necessary to review the story of industrial testing to see how the need for measuring dexterity gradually developed.

REVIEW OF RELATED STUDIES

Resume of industrial testing. Industrial testing has been in a laboratory experimental stage for about thirty years. Ever since Munsterberg[1] investigated several large manufacturing concerns, there has been a gradual acceptance of the value of evaluating the human factor in industry. His tests developed to select motormen for electric railways, ship pilots for ships, and girls for telephone operators, aroused much interest and suggested numerous possibilities for selection. However, there was another growing force in industry that helped prepare the ground for the development of scientific management pioneered by Taylor.

> Industrial psychology had profited not only from the adoption of the techniques of scientific management but from the demonstration by Taylor that industry can expect a definite return from an investigation and analysis of human behavior. It is partly because Taylor showed that returns to both the management and the work could be increased through the study of the human element that early psychologists were granted by industry an opportunity to conduct investigations designed to increase human effectiveness in industry. [2]

The growth of industrial psychology in the United States has been

---

[1] Munsterberg, H. Psychology and Industrial Efficiency, Houghton Mifflin, Boston, 1913, pp. VII, 3-320.

[2] Viteles, Morris S. Industrial Psychology. W. W. Norton Company, New York, 1932, p. 18.

the work of many research workers and teachers in the universities, or through research organizations which are supported largely by industrial firms.

Bingham's[1] report to the directors of the Personnel Journal December 1928, indicated a general trend of improvement in employment procedures. He commended a number of universities for their experiments in employment tests. Likewise, he mentioned the investigation of employment tests in the Dennison Manufacturing Company, the Scovill Manufacturing Company, the Eastman Kodac Company, and the Western Electric Company.

Pond's[2] investigation in the Scovill Company was one of the most comprehensive studies made to that time. In an effort to find if there were any relationship between factory worker's intelligence and his success and stability, it was reported "no marked agreement was found between earnings, foreman's ratings and length of service."

As we see, one of the most fundamental problems in the earlier experiments in measurement of vocational competency was how significant was general intelligence in fitting the person to the job. Cowderly[3] tested over 600 boys in a state trade school and reported favorable

---

1 Bingham, Walter V. "The Personnel Research Federation in 1928 Annual Report of the Directors", Personnel Journal, Vol. 7, pp. 299-313, December 1928.

2 Pond, Millicent, "Selective Placement of Metal Workers", Journal of Personnel Research, Vol. 5, pp. 346-368, January; Vol. 5, pp. 405-411, February; Vol. 5, 452-466, March, 1927.

3 Cowderly, K. A. "Measures of General Intelligence as Indices of Success in Trade Learning", Journal of Applied Psychology, Vol. 6, pp. 311-329, December, 1922.

results with trade success. Thurstone[1] reported general intelligence had no value in diagnosing ability to learn telegraphy. Other reports brought in similar results.

Experimental research showed the need for measuring other abilities rather than general intelligence to determine vocational fitness. It also brought a challenge for the need of a clearer definition and an agreement upon the concept of intelligence. The factors in vocational selection broaden into areas beyond intelligence. Mechanical comprehension, dexterity, manipulative ability, as well as personality and interest, were measured and evaluated.

Reports on studies of manipulative or dexterity tests. The different points of view as to the nature of mechanical ability has flooded the market with different types of tests. Viteles[2] noted four points of view that have influenced the development of a number of mechanical tests. The single factor point of view strongly supported by Stenquist and further advocated by Cox, influenced the development of a number of mechanical tests in an effort to measure this mechanical ability. In contrast to this point of view, Perrin and Seashore have been led to the conclusion of many specific skills rather than any theory of general motor ability. Other investigators supported the concept of the existence of group factors as a determiner of mechanical ability. This concept has been supported by such men as Farmer, Gaw, Spearman, and others.

---

[1] Thurstone, L. L. "A Study of the Diagnostic Value of Mental Test For Predicting Ability To Learn Telegraphy", Journal of Applied Psychology, Vol. 3, pp. 110-117, June 1919.

[2] Op. cit., pp. 231-238.

The fourth point of view, as suggested by Viteles, looks to the function of motor skill rather than its constituents parts.

One of the early finger dexterity tests was designed by O'Connor.[1] It is known as the Johnson O'Connor Finger Dexterity tests. He also constructed the Tweezer test. These two tests have been used extensively in industry. It has been reported favorable in some situations but generally not quite so favorable. Candee[2] and Blum, who made their study in a watch factory, reported that this test showed a difference between superior and mediocre workers. The correlation between foreman's ratings and test scores was .26.

Tiffin and Greenly,[3] who used the O'Connor Finger and Tweezer Dexterity in selecting employees for electrical fixtures assemblers and radio assemblers, reported a low correlation of .22 and .33 between earnings and merit ratings with the test scores.

Otis[4] reports a minus correlation, -.17, between successful power machine operators and the O'Connor dexterity test. However, in his study Otis noticed a difference in the operators with regard to speed and quality. He came to this conclusion:

---

[1] O'Connor, Johnson, Born That Way, Williams and Williams Company, 1929, pp. 1-323.

[2] Candee, B. and Blum, M. "Report of a Study Done in a Watch Factory", Journal of Applied Psychology, Vol 21, pp. 572-582, October 1937.

[3] Tiffin, Joseph, and Greenly, R. J. "Employee Selection Tests for Electrical Fixture Assemblers and Radio Assemblers", Journal of Applied Psychology, Vol. 23, pp. 240-263, April 1939.

[4] Otis, J. L. "The Prediction of Success in Power Sewing Machine Operating", Journal of Applied Psychology. Vol. 22, pp. 350-366, August 1938.

> There seems reason to believe that the use of independent batteries for predicting success on a specific job in terms of quality and speed of performance, respectively, is a novel approach in the study of occupational adjustment. Justification for this approach is to be found in the practical situation in industry where speed and quality are given different emphasis by different employers, and in certain psychological studies on the relationship between quality and speed of performance, which show that in many instances it is necessary to treat them as independent variables. [1]

Another test that is being used extensively in industry is the Minnesota Rate of Manipulation Test. This test was devised to measure native speed capacity. Ziegler[2] says, "Speed is primarily a native trait, improvable only at a very limited extent".

Blum[3] and Candee made a study in two large department stores that selected packers and wrappers by the use of the O'Connor Finger Dexterity and the Minnesota Placing and Turning tests. The multiple correlation for all three was only .38, just slightly higher than the Placing test alone.

Among other dexterity tests that are on the market are the Hayes Pegboard, Purdue Hand Precision, and Purdue Dexterity. Few studies have come in validating their success as measures for required skills.

The job analysis technique is still much used. Drake[4] has been an advocate for designing special tests from the job analysis and using time

---

[1] Ibid. p. 364.

[2] Ziegler, W. A. Manual of Minnesota Rate of Manipulation Test, Educational Publishers, Philadelphia, 1939, p. 1.

[3] Blum, M. and Candee, B. "The Selection of Department Store Packers and Wrappers with the Aid of Certain Psychological Test", Journal of Applied Psychology, Vol. 25, pp. 76-85. February 1941.

[4] Drake, C. A. "Aptitude Test Help you Hire", Factory Management and Maintenance, Vol. 95, pp. 55-57, June 1937.

and motion study techniques to furnish the basic data for the approach to right testing.

A recent report from the Woodward Governor Company[1] justified the battery of aptitude tests which was designed by the Psychological Corporation. No statistical reports were made on these special designed tests but only a favorable general report on the battery used.

SUMMARY OF RESEARCH

In this nebulous stage of manipulation or dexterity testing, a definite approach is being established to study the complexity of the job to find the basic skill elements required for successful performance before a test is selected or constructed. The simpler the job elements the simpler to construct a test for those elements.

The review of previous research seems to point out that:

1. There is a need for clearer concepts of the factors that go to make up measures for different abilities in determining vocational fitness.
2. The relation between dexterity test and success on job show very low positive correlations as found in the literature.
3. There is a great need for more refined procedures in evaluating dexterity tests.

PURPOSE OF THE STUDY

In view of the need of more critical investigations of dexterity

---

[1] Martin, I. "How and Why Woodward Governor Company Used Aptitude Tests for Employment", Sales Management, Vol. 48, pp. 18-20, March 1941.

tests, the Threading Sleeves Test will be analysed in an effort to find out how accurately the test may be used as an aid in selecting Mount Operators.

CHAPTER II

THE TEST AND THE TYPE OF WORK IN THE MOUNT DEPARTMENTS

DESCRIPTION OF TEST

The Threading Sleeves Test was constructed after careful examination and observation of the work performed in the mount departments. Since most of the work was assembling very small parts, this test was designed to use generally the same small muscles of the fingers used in grasping and manipulating mount parts.

Layout of materials. Figure 1 shows the layout and the materials used in this test. To the right on a blue sheet of paper, there is a pile of sleeves whose length is 5.5 millimeters with an inside diameter opening of .023 millimeters. To the left are several lead wires whose length is 78 millimeters and outside diameter .016 millimeter. Holding a lead wire in the left hand, if right handed, the subject, using the tweezers in the right hand, picks up one sleeve at a time and places it on the lead wire. This continues until ten sleeves are on the wire. Then other wires are used in the same manner for the number of trials allowed. Time is noted with a stop watch for each trial in threading ten sleeves on a wire.

INSTRUCTIONS FOR GIVING THE TEST

The subject is seated in front of a table 30 inches high. The materials on the blue paper are placed before him.

Figure I

Picture of Test Materials

Instructions. "This is called the Threading Sleeves Test. It is to see how quickly you can thread ten sleeves on a lead wire. These are the sleeves (point to pile of sleeves). They have a hole through them. You thread them with your tweezers onto the wire like this. (Show subject how to place one sleeve on the wire). Continue until you get ten sleeves on the wire. The wire will hold only about ten sleeves but be sure and count them. When you get ten on the wire say TEN."

Holding materials before test. "(Have subject pick up a lead wire.) Hold the wire with your left hand if you are right handed. (Check to see wire is held correctly. Wire should be held vertically near the top of wire.) Take tweezers with your right hand and hold like a pencil. When I say BEGIN you pick up a sleeve with your tweezers and start threading it on the wire. Remember, say TEN when you get ten sleeves on the wire. Get ready, BEGIN. (Note time with stop watch. Record each trial separately.)"

DESCRIPTION OF WORK IN MOUNT DEPARTMENTS

Mounting radio tubes. The work performed in the mount departments is assemblying very fine parts to make the mount or inside of a radio tube before the glass and base are attached. This work is assembled by operators in different positions on the team. The first operator or first position assembles a number of parts and passes the mount to the next operator where she in turn adds other parts to the mount, and so on down the line. The number of operators for each team may vary according to the difficulty and the amount of work that goes into a tube type. On some teams there may be only 2 operators, whereas on other teams, as many

as 8 or more operators may be required to complete the mounting of a tube. The pieces of materials, likewise, vary according to specifications of the tube types. Naturally, the amount of work or number of mounts assembled by teams whose specifications are different, would be affected according to the team line-up. The average number of mounts for some types would be approximately 30 mounts per hour for one team, where-as, a different tube type would average 60 mounts per hour for that same team.

This non-conformity of team members, parts, and amount of work performed, offer many difficulties for this study. However, to take the job of mounting as a whole, there is a general hand and finger manipulation that is constant and similar throughout the entire group of tube types. It was this general manipulation or dexterity of the hands and fingers involved in mounting that the Threading Sleeves Test used as a basis for construction of the test.

Mount line-up. Figure II shows the mount line-up for specific tube type which shows a typical sample of the division of work on a team. This line-up divides the team into 6 positions. The general movement of each position is picking up (P. U.) parts and assemblying these. In most cases tweezers are used either to pick up the parts or to aid in assemblying. The 1st and 2nd positions in this line-up will be broken down to see the similarity of manipulation on the job of mounting compared with that of threading sleeves on a wire as performed on the Threading Sleeves Test.

Figure II

MOUNT LINE-UP - TUBE TYPE 6SK7Gt

1st. & 2nd. Positions

P.U. & position mica cathode tab in jig
P.U. & assemble #1 grid
P.U. & assemble #2 grid
P.U. & assemble #3 grid
P.U. & assemble plate
P.U. & assemble top mica
Bend 3 plate tabs above top mica, make team mark
Remove unit from jig, bend 3 plate tabs over bottom mica, aside unit

Stock

3rd. Position

P.U. unit & collar. assemble collar to bottom mica
Bend 3 collar tabs over bottom mica, bend cathode ribbon. aside unit
P.U. stem & interlead shield. position & weld shield to #1 pin
P.U. position & weld connector to #2 stem pin
P.U. position & weld connector to #7 stem pin. aside stem

Stock

4th. Position

P.U. unit & connector, weld connector to #1 grid, bend connector
P.U. & weld connector to #2 grid, bend connector
P.U. & weld connector to #3 grid
P.U. stem & assemble to unit
Weld plate to #8 pin. weld mica eyelet to #3 pin. aside mount

Stock

5th. Position

P.U. mount. position, weld & adjust heater legs
Position & weld #2 grid connector to #6 pin
Position & weld #1 grid connector to #4 pin & #3 grid connector to #3 pin
P.U. position & weld connector to collar & #1 pin. aside mount

Heater Reoperation

Stock

6th. Position

P.U. mount & shield, assemble shield
Position mount in mandrel & weld (4) shield to collar. remove from mandrel
Position & weld cathode ribbon to #5 pin
P.U. position & weld getter to #3 grid leg
P.U. & assemble resilient mica
P.U. & assemble 2nd. resillent mica, aside mount
Inspection, count & tray handling

Stock

P.U. & position mica cathode tab in jig

The subject picks up and positions the mica cathode tab in the jig. The mica cathode tab is a thin flat disk with seven small holes arranged horizontally through it.

P.U. & assemble #1 grid

#1 grid is picked up with tweezers and assembled to the mica cathode tab. A grid is composed of wire wound around two posts leaving both ends clear.

The number of millimeters clearance depends on the specification of the tube type. The bottom grid legs are threaded or inserted into the two designated holes in the mica cathode tab. Tweezers are used to hold the top grid post to aid in guiding the bottom grid legs into the holes of the mica cathode tab.

P.U. & assemble #2 grid

#2 grid is picked up with tweezers and the grid legs are inserted in 2 of the holes of the mica cathode tab. Since #2 grid is smaller it is dropped between #1 grid.

P.U. & assemble #3 grid

In like manner the #3 grid is dropped between #1 and #2 grids and inserted into the mica.

P.U. & assemble plate

The plate is a small cylinder that fits around the grids. 3 tabs are on the bottom and the top of the plates. The bottom tabs are inserted in slots on the mica cathode tab.

P.U. & assemble top mica

The top mica is similar to the bottom mica. It too has 7 holes where

the 6 grid posts are inserted into the mica. The middle hole is used later to insert a heater which is a small thread-like wire. Tweezers are used to aid in directing the grid posts through the top mica.

<u>Bend 3 plate tabs above top mica, Make team mark</u>

Bending the plate tab over top mica helps to hold the mount in position. The tabs are bent with tweezers.

<u>Remove unit from jig. Bend 3 plate tabs over bottom</u>

The unit is removed from the jig and the 3 plate tabs are bent with tweezers to hold the mount. This completes the assembly of the unit for the 1st and 2nd positions.

The other positions show a similar breakdown of work. Note the general movement is picking up very small pieces with tweezers and placing them next to another small piece to be welded or inserted. The timing of the work has been divided by job analysis and motion studies. This gives a uniform amount of work by different positions on the same tube type.

In comparing the Threading Sleeves Test with the work performed in mounting there are several conditions that delimits this study. These limitations will be noted in the discussion of acquiring data for the analysis of the test.

CHAPTER III

PROCEDURES AND STATISTICAL TECHNIQUES

SCOPE OF PROBLEM

The subjects used in this study were a typical group of working women that were employed in the mount departments of an industrial plant. The tests were administered to the employees under controlled conditions. The influx of war workers increased instability among employees and accelerated the turn over problem. In many situations the employees were not in the department long enough to be rated by supervisors or trained on the job sufficiently to acquire production records from which criterion data could be obtained. This naturally presented a perplexing problem in acquiring sufficient data for statistical analysis.

To determine the value of the Threading Sleeves Test, the reliability and validity must be ascertained. Viteles[1] gives the following definitions or reliability and validity:

> Reliability - One measure of a satisfactory test, and of every other selection device, is its reliability. To be reliable a test or battery of tests must be a consistent measure of the relative ability of the members of a group.
>
> Validity - From the viewpoint of ultimate usefulness, one of the most important steps in the research program for the development of tests is the determination of whether the test or battery of tests actually taps

[1] Op. cit., pp.249 and 255.

> qualities of importance in success on the job, and whether it can be used in predicting such success. This envolves a comparison between test score and proficiency on the job as represented in the criteria.

The reliability and validity of the Threading Sleeves Test will be considered by studying the following:

<u>Reliability or Consistency</u>

Retest of the Threading Sleeves for 107 employees

Part-Whole correlation of the Threading Sleeves Test on 115 employees

<u>Validity Criteria</u>

Production Records of employees

Ratings of employees by supervisors

Before the discussion of how these data were procured, it is necessary to consider the question of what score to use from the Threading Sleeves Test as a basis for the statistical analysis.

<u>What score to use from the test?</u> The Threading Sleeves Test called for threading 10 sleeves onto a lead wire. This was repeated 3 times and the time noted in seconds separately for each trial. The score could be either the total time for 3 trials or the best trial out of the three. The mean scores for each trial with the total and best trial are listed in Table I.

TABLE I

MEAN SCORES OF THE DIFFERENT TRIALS IN THE THREADING SLEEVES TEST

Trials	Means	S.D.	N
1st	75.96	28.5	115
2nd	70.40	25.1	115
3rd	66.83	27.5	115
Best	59.92	17.9	115
Total	216.50	67.4	115

It is noted that each trial shows an improvement in seconds. That is, the smaller the number given in seconds the better the score. To note the relationship between each of these trials see the correlations given in Table II.

TABLE II

CORRELATIONS BETWEEN DIFFERENT TRIALS IN THE THREADING SLEEVES TEST

	r
1st Trial with 2nd Trial	.598
2nd Trial with 3rd Trial	.578
Best Trial with Total Trials	.874

The high positive correlation of .874 between the Best Trial and the Total Trials indicates there is a very close relationship between these two scores. On the basis of this relationship, it was decided to use the Best Trial as the score for the Threading Sleeves Test.

PROCEDURE FOR ESTABLISHING RELIABILITY

Administering the retest. In order to secure a representative group of employees for the retest, one department was chosen. Arrangements were made with the department head for all the employees to retake the test on the day designated. No employee was made to feel she had to take the retest, nor was she made to feel the scores were to be used against her. Since it took less than five to ten minutes to administer the test, it offered an extra relief period for the operators. The entire department of one hundred and seven employees took the test. An unusually fine spirit of cooperation prevailed throughout the department. Instructions for giving the test were the same as for the first test as explained in Chapter II.

Table III shows the ability distribution of the subjects within the department that took the retest. The distribution was made on an eleven point scale ranging from very, very low L- to very, very high T+. The norms of all the tests in the battery were built on this eleven point scale using the quartile deviation as basis for the division. The general intelligence distribution is an average score from the Otis Quick-Scoring Mental Ability Tests, Wilson Six-Minutes Test on Alertness, and the Healey Form Board, Puzzle B. The total index is the total test scores for all the scores made on the tests in the battery.

TABLE III

ABILITY DISTRIBUTION OF SUBJECTS TAKING THE THREADING SLEEVES RETEST

Scale[a]	General Intelligence	Total Index
T+		
T		
T-		
Q1+	5	2
Q1	22	16
M	46	50
Q3	20	17
Q3-	4	12
L+	9	5
L	1	4
L-		1
N	107	107

[a] This scale was built on the Quartile Deviation basis. M represents the median score while 25% cases above and 25% cases below locates the Q1 and Q3 points. The difference between the scores from M to Q1 and from M to Q3 divided by 2 gives the semi-inter quartile range or the Q distance. The remaining scores above Q1 and below Q3 are divided by the Q basis on these points indicated on the scale. The range for each point is smoothed out according to the spread of the scores. The norms for all the tests given in the battery used this scale.

It is apparent from this table that the distributions are slightly skewed in the direction of poor ability. This deviation probably represents a trend which would be typical of all the workers after being in the plant over a period of time. Usually the high ranking groups either become dissatisfied and quit the job, or are promoted to other jobs.

There was one factor that was not controlled within this group that probably would affect this study on retesting. That is, the operators had an uneven number of days of work experience on the job. In some cases operators had worked in the plant over 400 days, whereas others, had only 20 days experience working on the job. For a check on this variable factor the distribution of the Threading Sleeves scores of the 1st test and of the retest were made on the groups having more than 200 days work experience and less than 200 days work experience.

TABLE IV

DISTRIBUTION OF FIRST TEST AND RETEST OF THE THREADING SLEEVES TEST ACCORDING TO NUMBER OF DAYS WORKED

	More than 200 days		Less than 200 days		Total Work Experience	
Scale	1st Test	Re-Test	1st Test	Re-Test	1st Test	Re-Test
T+						
T						
T-		1		1		2
Q1+		6	7	7	7	13
Q1	12	18	9	22	21	40
M	13	9	18	10	31	19
Q3	7	4	8	1	15	5
Q3-	7	3	1	5	8	8
L+	2	1	2	1	4	2
L			5	2	5	2
L-	1		2	3	3	3
N	42	42	52	52	94	94
Median	M	Q1	M	Q1	M	Q1

Table IV shows probably there is no great difference in the medians of the groups that worked over 200 days and less than 200 days. The medians fell for both groups at M for the first test and Q1 for re-

test. However, the group that had less work experience shows a wider range of scores on the scale than the experienced group. But since the retest for both groups indicates a slight improvement, as the distribution shows, probably the factor of uneven days of work would not affect the statistical analysis for the retest to any great extent to justify it as a factor in affecting the tests reliability unless a more refined analysis were made.

Part-Whole correlation of Threading Sleeves Test. The Threading Sleeves Test is a delicate and intricate instrument with many variables affecting the timing of the test. Such variables as nervousness, error of exact timing in seconds, and others, make it difficult to establish the reliability of the Threading Sleeves Test. To compute a reliability coefficient for a Part-Whole correlation, 115 cards were taken from the files at random. The Best Trial out of the three trials were correlated with the Total Trials. The results of this correlation is shown in the chapter on the Analysis of the Data.

CRITERION DATA FOR VALIDITY AND PROCUREMENT

What criteria to use? Criteria of vocational success is usually divided into two groups - objective and subjective materials. Viteles[1] differentiates those as factors which can be expressed as objective terms in varied standards of accomplishment, whereas, a subjective criterion is submitted in the form of ratings of accomplishment. Efforts were made to get types of criteria as, production records for the objective criteria, and ratings from supervisors for the more subjective criteria. The question to be considered with this material is how well

[1] *Op. cit.*, p. 204.

does the good scores on the Threading Sleeves Test distinguish between the good and poor groups of employees.

Production records. Production records of 40 mount operators were used for a comparative study with the Threading Sleeves Test. The number 40 is a small sampling of the employees working in the mount departments. However, due to the fact that production records were made on teams rather than on individual scores, it became necessary to plan an experimental group in order to acquire individual records. The department chosen was a group of 8 teams all working on the same tube type. Each team had 5 operators. With the cooperation of the foreman and supervisors, provisions were made in order that each operator could work to her speed capacity for one day without waiting for materials from the team member next below her. That is, an inventory of work was made ahead for each position, so that, if the work was not fed down the line fast enough, the inventory would be used. An hourly count was made by the supervisor for each operator during the day. Only one team a day made the experiment as this made it easier to control several factors. This likewise gave the writer an opportunity to observe the work of the operators. In most cases they seemed to work to their capacity. A careful record was made of these observations.

The question arose as to what score would be the most reliable to use for comparison with the test scores. It was noted that 33 out of the 40 operators ranked the same on the total number of pieces for the day and their best hourly count for the day. That is, if an operator made the highest number of mounts on the team for the day, her best hour count was highest. The operators making the second highest on

total number of mounts made second highest on hour count for the team. This held true for 33 of the 40 operators. Observation of the day's work showed several uncontrolled factors, such as, machine trouble, and other time-out elements, that would likely be added in the total day's work, which would probably not be influenced if the best hour's count were used. Since these two counts gave approximately the same results, the best hour counts was the score selected for use.

To see how the operators in this experiment were a representative group as measured by the battery of tests, their total indexes of all the tests are distributed in Table V.

TABLE V

ABILITY DISTRIBUTION OF THE PRODUCTION RECORD GROUP

Scale	Total Index
T+	
T	
T-	1
Q1+	6
Q1	9
M	18
Q3	4
Q3-	1
L+	1
L	
L-	
N	40

This group seems to show a fairly normal spread of ability as measured by the battery of tests. This group may be said to be fairly typical of other mount workers.

<u>Ratings of employees</u>. Ratings are known to have many defects. Nevertheless, for this study, they furnish the most reliable criteria for success on the job. During the early ratings by supervisors, some of the major defects were noted, such as:

1. Ratings were influenced by the "halo" effect.
2. It was difficult for the supervisors to separate ability to perform the job from the attitude of the workers.
3. They tend to use only a few ratings.

To overcome some of these defects, the following attempts were made to construct a rating scale that would be more reliable in getting actual efficiency scores.

1. Encourage supervisors to discriminate between factors that made for good or poor workers, such as skill, production, quality, and speed.
2. Two ratings were required on separate charts. One for ability to perform the job, the other, the attitude of the workers.
3. To eliminate the tendency to use few ratings, a rating scale was constructed using the concept of the normal distribution curve. A certain percentage of the group was encouraged to be placed in a particular section under the curve.

The rating scale was not fully constructed and used by supervisors and foremen until the need for interpreting the test scores became evident. This became a slow process of guidance and learning, but gradually the concept of the normal distribution curve was understood,

as they learned the meaning of spread of different ability. The early scale used terms in rating the operators, such as, exceptional, good, average, fair, and poor. The letters T, Q1, M, Q3, and L were applied to these terms. Later this rating scale was enlarged to use the same 11 points scale on which the norms of the tests had been constructed. A large scale was drawn on a sheet of paper 15 inches by 18 inches, and divided into eleven divisions. The names of the operators were placed under the curve accordingly as to their abilities to perform the job. Numerical values were given to each division for statistical analysis. A sample of the rating sheet is drawn on a smaller scale and is found in Appendix 1.

The following ratings were secured for this study.

Ratings	Department	N
Foreman's Ratings	one	47
Foreman's Ratings	two	107
Supervisor's Ratings	one	53
Supervisors' Ratings	five	255

The foreman's ratings from one department and the supervisor's ratings from one department were the same group rated by each. The ratings from five departments were secured from five different supervisors in the mount departments within the plant. The supervisors used the early scale for ratings as they were not acquainted with the test interpretations; but, the foreman used the 11 point scale for rating. There is a possibility that in the foreman's ratings a knowledge of operator's test scores may have entered into the ratings, yet, no doubt, the knowledge of better how to evaluate the work of the operators also entered into the ratings.

STATISTICAL TECHNIQUES

Reliability and validity are expressed in terms of coefficient correlation. The Pearson product-moment coefficient correlation is the statistical technique used in this study to measure the relationships. Data for the following were thus studied:

1. Retest of Threading Sleeves Test vs Threading Sleeves Test
2. Best Trial vs Total Trials of Threading Sleeves Test
3. Production Records vs Threading Sleeves Test
4. Supervisors' Ratings vs Threading Sleeves Test
5. Foreman's Ratings vs Threading Sleeves Test
6. Tests in the battery vs Threading Sleeves Test

CHAPTER IV

ANALYSIS OF DATA

Treatment of data. The purpose of this study was to determine the reliability and validity of the Threading Sleeves Test in order to find out how accurately it could be used as an aid in selecting employees with the potential to become good operators. In chapter III detail description of the criterion data and the steps in procurement were discussed. This chapter summarizes the results of the findings from the analysis of these data. The following questions are to be considered.

1. How reliable is the Threading Sleeves Test as shown by the retest?
2. Does the Part-Whole correlation of the Threading Sleeves Test indicate that the test is a consistent measuring instrument?
3. What relationship is indicated between the test and the production records?
4. How well does the good scores on the Threading Sleeves distinguish between the good and poor groups as rated by the supervisors?

Coefficient Correlations of Reliability

How reliable is the test? This question is answered by considering two correlations; first, a retest correlated with the Threading

Sleeves; second, a Part-Whole correlation of the Threading Sleeves. Table IV shows the results of the test's consistance from these correlations.

TABLE VI

RELIABILITY OF THE THREADING SLEEVES TEST

	r	P.E.	N
Threading Sleeves vs Retest	.568	±.044	107
Part vs Whole of test	.874	±.015	115

The coefficient reliability of .568 is not sufficiently high to place too much confidence in the consistancy of this test. Apparently other factors that had an influence on the test were not involved in both tests. Even though the materials and instructions were the same, the conditions under which the subjects took the tests were somewhat different. During the first test when the subjects were being hired, they were more nervous and fumbled with the materials, but during the retest, after having experience working with small parts, the subjects were less nervous and handled the materials with more assurance.

There was no way that this factor could be measured or controlled, therefore no definite statement of it's influence on the test can be made other than this general statement based on observation of it's probable affect.

The Part-Whole correlation of .874 offers a higher coefficient of reliability than the retest. This measure was reported in Table II when the question arose as to what score to use for the Threading

Sleeves Test. Three trials were given to thread 10 sleeves on a wire. Each trial was timed separately. In deciding what score to use, either the best trial out of three trials or a total of the three trials, the high correlation between these two were noted which showed that either score would give approximately the same results. As this correlation offered a consistant relationship, which represented a part and the whole of the test, it was again reported here to show the consistancy of the test. This reliability coefficient offers a measure that would justify the use of the Threading Sleeves Test as a reliable instrument.

Validity Data

The validity of the Threading Sleeves Test is shown by the relationship between scores on the test and success on the job as indicated by the results of the following correlation table.

TABLE VII

RATINGS AND PRODUCTION RECORDS CORRELATED WITH THREADING SLEEVES TEST SCORES

Threading Sleeves vs -	r	P.E.	N
Production Records	.134	±.064	34
Supervisor's Ratings	.159	±.087	57
Supervisors' Ratings	.211	±.013	255
Foreman's Ratings	.364	±.085	47
Foreman's Ratings	.344	±.056	107

These correlations are generally low. In some cases the numbers

are so few they are not statistically significent as indicated by their P.E. scores. This is particularly true for the Production Record correlation of .134 and the first Supervisor's Ratings correlation of .159 which shows their P.E. scores obtained no statistical measure that could not be due merely to the fluctuations of random sampling.

TABLE VIII

CORRELATIONS OF PRODUCTION RECORDS WITH OTHER TESTS USED IN THE PROPOSED SELECTION FOR MOUNT OPERATORS

Production Records vs -	r	P.E.	N
Otis Mental Ability	.021	$\pm$ .115	34
Wilson Six-Minutes Test of Alert.	.074	$\pm$ .111	36
Healey-Fernald Const. Puzzle B.	.030	$\pm$ .112	36
Stenquist Mechanical Ability	-.291	$\pm$ .102	36
Witmer's Cylinder Test	.488	$\pm$ .085	36
Threading Sleeves Test	.134	$\pm$ .064	34

The first three correlations, the Otis Mental Ability .021, the Wilson Six-Minutes Test of Alertness .074, and the Healey-Fernald Construction Puzzle B .030, show practically no relationship with the production records and general mental ability as measured by these tests. The Stenquist Mechanical Ability shows a negative relationship of -.291. However, the Witmer's Cylinder Test, which is a manipulative test, indicates a higher relation with production records as shown by the correlation of .488, than the Threading Sleeves Test with the correlation of .134.

TABLE IX

COMPARISON OF CORRELATIONS FROM THE RATINGS FOR THE FIRST AND RETEST OF THE THREADING SLEEVES TEST

Ratings vs -	r	P.E.	N
Supervisor's Ratings vs First Test of Thr. Sl.	.159	±.089	57
Supervisor's Ratings vs Retest of Thr. Sl.	.260	±.079	62
Foreman's Ratings vs First Test of Thr. Sl.	.364	±.085	47
Foreman's Ratings vs Retest of Thr. Sl.	.604	±.058	53

This table shows what relation the same ratings have with the first test and the retest. In both cases the correlations are higher for the retest. The foreman's ratings with the retest of .604 implies a very marked correlation and is to be regarded as high and somewhat useful when compared to the best findings in the industrial testing field.

TABLE X

CORRELATION OF THREADING SLEEVES WITH OTHER TESTS IN BATTERY

Threading Sleeves vs -	r	P.E.	N
Otis Test of Mental Ability	.074	±.035	355
Wilson Six-Min. Test of Alertness	.137	±.035	352
Healey-Fernald Const. Puzzle B.	.021	±.036	352
Stenquist Mechanical Ability Test	.065	±.036	341
Minnesota Rate of Manip., Turning	.339	±.100	76

These correlations are generally low as would be expected. In selecting tests for a battery different types of tests are chosen in order to measure different aspects of an individual. Apparently the Threading Sleeves Test must be measuring some other elements or factors than the tests listed above are measuring as indicated by the low correlations.

It is noted that practically all the validity criterion data submitted for this study signified positive yet low correlations. Nevertheless, when these are compared with the findings of other research in the field of dexterity testing, these correlations appear more significant. Those reported in Chapter I on the "Review of Related Studies" are listed again below.

O'Connor Finger and Tweezer Dexterity correlated with ratings from a watch factory	.26
O'Connor Finger and Tweezer Dexterity correlated with earning and merit ratings from radio assemblers	.22 and .33
Multiple correlations of Minnesota Placing and Turning Test and O'Connor Finger Dexterity correlated with large department store for selecting packers and wrappers.	.38

Other favorable reports of dexterity tests are made but they are not substantiated by statistical data. Tiffin[1] points out, "no single test will measure all of the capacities or abilities required on any job". In the light of this it may be noted that the Threading Sleeves Test offers as satisfactory a single test as are generally reported. In actual practice, in the factory where developed, the Threading Sleeves

---

[1] Tiffin, Joseph, Industrial Psychology, Prentice-Hall, New York, 1944, p.42.

was used, not alone, but as one of a battery of tests. Supervisors were taught to avoid reliance on a single test, but instead, to study individual's score on a battery of tests, as an interlocking pattern.

CHAPTER V

SUMMARY AND CONCLUSIONS

SUMMARY

The Threading Sleeves Test is a dexterity test that was constructed for a specific job in an industrial plant where most of their applicants were placed. The purpose of this study is to ascertain to what degree the Threading Sleeves Test was a reliable measure to be used in selecting mount operators for the radio tube factory. The type of work performed on this job was a repetitive job of assemblying very small parts to construct the inside of a radio tube. This was performed by operators on a team. Since the number of operators and amount of work performed by each varied in accordance with the specifications of tube types, this increased the difficulty of acquiring data to analyse the validity of this test. Briefly stated the reliability and validity of this test were considered by studying the following:

1. Retest of the Threading Sleeves Test
2. Part-Whole correlation of Threading Sleeves
3. Production records of employees compared with Threading Sleeves Test.
4. Ratings by supervisors and foremen compared with Threading Sleeves Test.
5. Review of literature as reported on dexterity testing performed in other industries.

CONCLUSIONS

The conclusions based on the evidence acquired must be considered with reserve. The groups are predominately small, the date complex, and the factors related to job success are many and varied.

1. The reliability coefficient based on a retest of 107 after an interval of 20 days to 400 days of work experience was .568. This is not considered a high correlation for a retest, although encouraging. The degree of nervousness was not constant during both tests as was observed.
2. The reliability coefficient of .874 found by the Part-Whole method, that is, the best trial out of three, correlated with total of the three, was considered sufficiently high to justify the use of the best single performance of the test as a reliable measurement score.
3. The production records of 34 operators with their scores on the Threading Sleeves Test gave correlations so low as to be statistically non significant.
4. In this study, it was found that all ratings showed a positive relationship between good and poor workers with high and low scores on the test. Correlations of .211, .344, .364, and .604 indicated the relationship between the Threading Sleeves Test and the ratings of workers by their supervisors or foremen.

5. In comparing the validity of this test with results of reported studies found in the literature, this study indicates the Threading Sleeves Test as shown by coefficient correlations are as high and in one instant higher as are generally reported.

SIGNIFICANCE OF THIS TEST

The significence of this test has already been indicated in the discussion of validity and the above conclusions. While high scores on this test are positively related to success on the job that calls for speed and skill in finger manipulation, it is both unwise and impossible to expect one test to select operators on the basis of just the test scores alone. As an aid for selecting, it could be used as a supplementary scale with reasonable confidence. The description and job demands vary so greatly among jobs and companies that this test is only significant to the particular job of assembling radio tubes.

FURTHER RESEARCH

1. Further research is needed on the reliability of the test. Administer a retest on the following day after the first test rather than having twenty to four hundred days of experience on the job between administering of the second test as reported in this study.
2. Research on the influence of practice on the job; is improvement shown adequately measured by the test.
3. Give the Threading Sleeves to other industrial situations

which require finger dexterity, in order to further study the help that this test may offer.

4. Correlate the Threading Sleeves Test with other finger and tweezer dexterity tests that have been standardized.
5. Report the standardization of the Threading Sleeves Test.

BIBLIOGRAPHY

BIBLIOGRAPHY

Bingham, Walter V. "The Personnel Research Federation in 1928, Annual Report of the Directors", Personnel Journal, Vol. 7, pp. 299-313, December 1928.

Blum, M. and Candee, B. "The Selection of Department Store Packers and Wrappers with Aid of Certain Psychological Tests", Journal of Applied Psychology, Vol. 25, pp. 76-85, February 1941.

Candee, B. and Blum, M. "Report of a Study Done in a Watch Factory", Journal of Applied Psychology, Vol. 21, pp. 572-582, October 1937.

Cowderly, K. A. "Measures of General Intelligence as Indices of Success in Trade Learning", Journal of Applied Psychology, Vol. 6, pp. 311-329, December 1922.

Drake, C. A. "Aptitude Test Help You Hire", Factory Management and Maintenance, Vol. 95, pp. 55-57, June 1937.

Martin, I. "How and Why Woodward Governor Company Used Aptitude Tests for Employment", Sales Management, Vol. 48, pp. 18-20, March 1941.

Musterberg, H. Psychology and Industrial Efficiency, Houghton Mifflin, Boston, 1913, pp. VII, 3-320.

O'Connor, Johnson, Born That Way, Williams and Williams Company, 1929, pp. 323.

Otis, J. L. "The Prediction of Success in Power Sewing Machine Operating", Journal of Applied Psychology, Vol. 22, pp. 350-366, August 1938.

Pond, Millicent, "Selective Placement of Metal Workers", Journal of Personnel Research, Vol. 5, pp. 346-368, January, - Vol. 5, pp. 405-411, February - Vol. 5, pp. 452-466, March, 1927.

Remmers, H. H. and Gage, N.L. Educational Management and Evaluation, Harpers Brothers, New York, 1934, p. 327.

Thurstone, L.L. "A Study of the Diagnostic Value of Mental Test for Predicting Ability To Learn Telegraphy", Journal of Applied Psychology, Vol. 3, pp. 110-117, June 1919.

Tiffin, Joseph, Industrial Psychology, Prentice-Hall, New York, 1944, p. 42.

Tiffin, Joseph, and Greenly, R. J. "Employee Selection Tests for Electrical Fixture Assemblers and Radio Assemblers", Journal of Applied Psychology, Vol. 23, pp. 240-263, April, 1939.

Viteles, Morris S. Industrial Psychology, W. W. Norton Company, New York, 1932, p. 18.

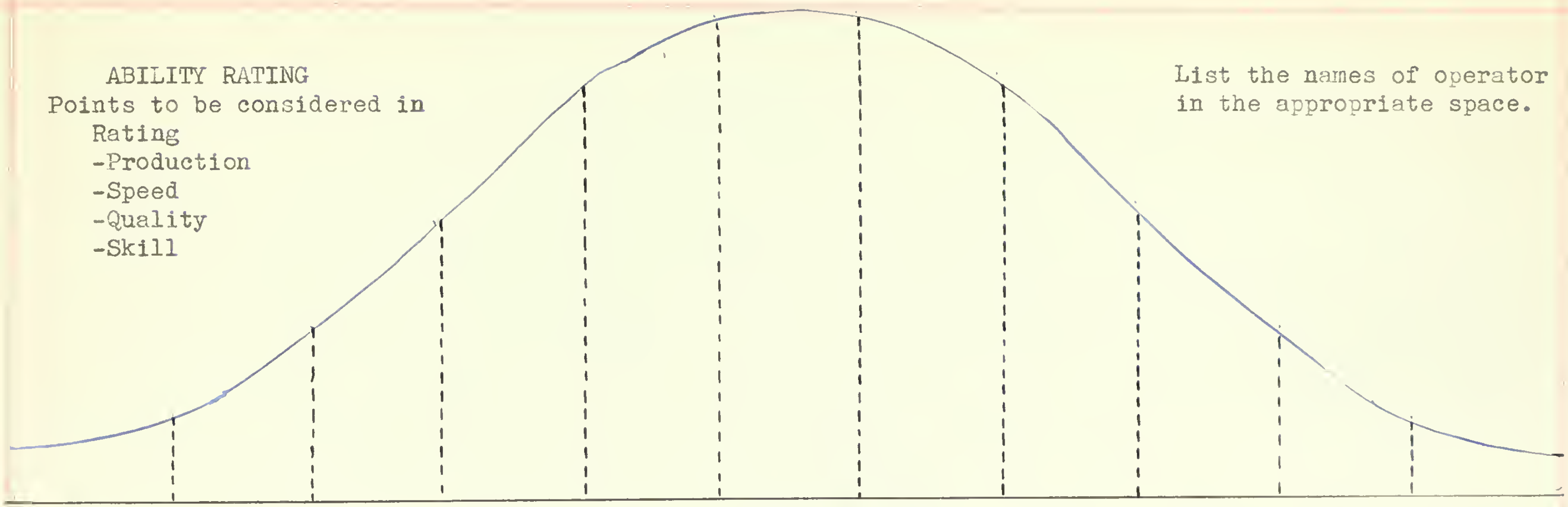
Ziegler, W. A. Manual of Minnesota Rate of Manipulation Test, Educational Publishers, Philadelphia, 1939, p. 1.

APPENDIX I

ABILITY RATING CHART

ABILITY RATING
Points to be considered in
Rating
-Production
-Speed
-Quality
-Skill

List the names of operator in the appropriate space.



L-	L	L+	Q3-	Q3	M	Q1	Q1+	T-	T	T+
Im-possible	Very Very Poor	Very Poor	Poor	Below Average	Average	Good	Very Good	Very Very Good	Superior	Very Superior

APPENDIX II

THE BATTERY OF TESTS

APPENDIX II

THE BATTERY OF TESTS

The aim of the testing program was to measure two phases of an individual's ability, viz., 1. mental alertness, and 2. manipulative ability, with emphasis on the manipulative of small parts.

Tests used to test mental alertness were: 1. Otis Test of Mental Ability; 2. The Wilson Six-Minute Test of Alertness; 3. The Army Beta Test; 4. Schooling of the individual; 5. The Healey-Fernald Construction Puzzle B. Correlations among these tests were reasonably high. The results were supplementary and re-enforcing.

Tests used to check on manipulative ability were: 1. The Stenquist Mechanical Aptitude Tests; 2. Minnesota Rate of Manipulation Tests, - two parts, placing and turning tests; 3. The Threading Sleeves Test; 4. The G. E. Stem Head Test; and 5. The Pennsylvania Bi-Manual Test. - two parts.

Seldom were all of these tests given to an individual. The battery changed from time to time, but usually there were three mental alertness tests and three manipulative tests, in the battery given any individual. For specific jobs there were other tests.

The Threading Sleeves Test was always used as one of a battery of tests, having its weight in a decision on ability, but never urges as the sole determiner of ability, not even manipulative ability.